Aspects of modality in audio-visual processes

de Gelder, B.; Bertelson, P.; Vroomen, J.

Published in:
Speechreading by humans and machines

Publication date:
1996

De Gelder, B., Bertelson, P., & Vroomen, J. (1996). Aspects of modality in audio-visual processes. In D. G. Stork, and M. E. Hennecke (Eds.), Speechreading by humans and machines. NATO ASI Series F, (vol. 150, pp. 179-192). Berlin: Springer-Verlag, Gmbh.

Aspects of modality in audio-visual processes

1 Beatrice de Gelder, Paul Bertelson and 2 Jean Vroomen
1 Cognitive Psychology and Psychophysiology unit, Tilburg University, The Netherlands, and Laboratoire de Psychologie Experimentale, Universite Libre de Bruxelles, Belgium
2 Cognitive Psychology and Psychophysiology unit, Tilburg University, The Netherlands

ABSTRACT

The relevance of the sensory modality in which information processing is initiated is not an issue that has received much attention in cognitive psychology. Much in the spirit of information processing approaches to cognition, psychological content is conceived as abstract. Mathematical approaches to information processing have likewise not made much room for conceptualizing the specific contribution of the sensory input modalities. Bi-modal speech processing is among the best explored cases, yet it is not clear that investigations of speech processing in another modality than the canonical auditory one, i.e., lipreading, have so far departed from the mainstream information processing frameworks. A more recent theoretical proposal which casts information processing as being a matter of separate modules or of separate content domains leaves room for modality specificity. Against this background two apparently related phenomena are discussed, McGurk effects and ventriloquism. Starting from models that neither postulate modality specific processes nor modality specific representations, the paper asks whether a distinction is needed between content driven and contiguity driven processes.

Many natural events consist of co-occurring or correlated occurrences in different sensorial modalities. Studies of perception investigate vision, audition, and the like mostly by focusing on the abstract content that is processed rather than on the specific contribution from the sensory modality itself. In line with an information processing approach to cognitive processes, cognitive content is conceived as abstract. Not surprisingly, approaches to information processing that are driven by mathematical formalisms have not made much room for conceptualizing the specificity or possible implications of the sensory modalities in the course of information processing. Bi-modal input situations have received less attention than has been devoted to the study of information processing in single modalities.

Among the cases of multi-modal input that have been studied systematically, bi-modal speech perception is a prime one. Yet it is not clear that investigations of speech processing in another modality than the canonical auditory one depart from approaches to information processing developed in the course of understanding single modality perceptual processes. In this paper we compare two phenomena based on bimodal processes and raise the question of the specific representations involved in them. We start by clarifying the notions of modality and modularity. Next, we present brief overviews of audio-visual speech as in the McGurk effect and of audio-visual pairings as in ventriloquism. We consider whether one might be a specific case of the other and vice versa. Finally, we propose to clarify the issues at stake by providing arguments against a modality neutral view of cross-modal integration and argue in favor of a distinction between content driven and contiguity driven processes.

I. Audio-visual speech and audio-visual pairings

Two audio-visual phenomena that appear to share important similarities are the McGurk effect and ventriloquism. These similarities raise the question whether the McGurk effect in speech perception might be a subspecies of more general situations of pairing auditory and visual information as in ventriloquism. Two questions dominate our concern with modality-specificity: the question of the specificity of representations and that of the specificity of processes. We will argue

that the present evidence seems to require the existence of modality-specific representations operating in a modular context. We will thus contrast McGurk effects with other kinds of spatio-temporal integration processes, like ventriloquism. The phenomena we discuss as well as the models offered for them can be organised along two contrasting dynamics leading to cross-modal integration which will be referred to as content-based and contiguity-based dynamics. The former would correspond to a set of principles and constraints, a mechanism for short, that puts together inputs that belong to a same content domain, irrespective of the modality in which the information is presented. The other would be a mechanism that is based on contiguity of two events in time and space or is otherwise based on Gestalt-like principles, irrespective of the actual content of the events.

1. Foundational and epistemological issues

Philosophers have traditionally been interested in modality-specificity of the empirical sources of our knowledge. The great debate turned around the question whether the senses must be considered as the ultimate anchors of objective knowledge or as sources of its contamination. Berkeley is a notorious protagonist in this debate with his claim that our knowledge begins and ends with experiences locked into sensorial modality-specificity. But in the context of present day empirical psychological theories, such concerns are hard to trace. The notion of modality-specificity that is of most immediate concern in psychological research relates to the claim about sense-specific coding of information, typically the notion that vision provides visual information, hearing provides auditory information, etcetera. But the issue is not about sensory knowledge of objects, nor about sensory concepts, but only about sensory-specific featural information. Information from the various modalities gets integrated and we achieve objective knowledge and entertain concepts, etcetera. Depending on one's philosophical and psychological preferences, one may hold the view that modality-specificity percolates upward the information processing system all the way leading to a concept of e.g., a visual circle which is different from that of a haptic circle. The alternative view is that modality-specificity begins and ends with the senses, that the impact of input modality is limited to how impressions are received and not to their content. Information or

content itself is a matter of abstract concepts and propositional knowledge including knowledge about sensory objects.

Although these issues are basically philosophical ones, they have at times played an important role in most noteworthy developmental psychology (Spelke, 1987). Philosophers have been attracted by developmental data showing early cross-modal processes and advanced these in support of epistemological views (Campbell, 1994).

2. An autonomous level of functional processes: Modularity

The long standing tension in philosophy between modality-specific sensory representations and abstract concepts has received comparatively little attention since the beginning of experimental psychology. With few exceptions, the intellectual context was one where the distinction between concrete modality specific input and abstract representations was taken for granted. It remains a common view in the literature on intermodal relations that the relevant contrast is the one between concrete, in the sense of modality-based inputs and abstract representations. Earlier research for example by Bertelson on ventriloquism was very much in the tradition of contrasting sensory and conceptual processes, perceptions and cognitions in the accepted sense of these terms. For example, experiments were set up to examine conceptual influence on sensorial integration where subjects' conceptual knowledge of what loudspeakers were for was the critical variable. If the notion of a mental module is useful, it hangs on the conceptual consistency and the empirical plausibility of a modular level of processing that is neither captured by sensory analysis nor by conceptual labour.

In recent developments of cognitive psychology and the neurosciences, attention has been drawn to what is now loosely referred to as the modular organisation of mind. The notion of a functionally autonomous competence and its corresponding processor has been argued for since two decades by neuroscientists as well as linguists in the Chomskyan tradition. Fodor (1983) summarised some main characteristics of the assumed properties of functional modules and has contributed to making the term more widespread among psychologists. The notion of a level intermediate between sensorial processes and central abstract thought

processes represents an intriguing development and a challenge for understanding the locus of modality specificity. What concerns us most critically here is that the existence of such a level would seem to call for a revision of the traditional contrast between sensorial and abstract cognitive representations.

So far very little attention has been paid to the way sensorial modality and mental modularity on the one hand and modularity and abstract cognitive representations might be combined. For example, Massaro (1987) argued that facts which transcend modality-specificity of speech present arguments against the modularity of mind. The notion underlying this view is that information is specific, but that the processes operating on it are always the same (Massaro, this volume). Against this approach, we have argued before that modality and modularity are orthogonal issues (de Gelder and Vroomen, 1989) and we pursue that argument here.

How might a framework where three different levels of processing, sensorial, functional/modular, and cognitive are distinguished, change and advance our ways of thinking about modality? Modality and modularity are complex and ambiguous notions. This ambiguity derives from the fact that the two occur up at various stages in psychological models of knowledge and in philosophical analyses of its foundations. Yet, the two notions can be used unambiguously when the nature of the explanation they figure in is made explicit. In what follows, we will rely on a specific meaning of modularity that we believe to be essential one for the issues at stake in understanding audiovisual processes.

A central issue is to understand how matters of modality of the senses are very different from issues of the modularity of mind. The distinctive characteristic of a module is its domain-specificity, and it rests on the claim that a module is a processing mechanism that is operational in a certain domain of phenomena. Typical examples are the class of speech events, faces, etcetera. In contrast, claims about modality relate to questions on the sensorial mode of information input and correspond to ways of carving information input up into sensorial regions. Claims about modularity carve information up into categories of objects, or more specifically, into semantic domains. Besides being contrasted with sensory processes, modular processing is also different from the kind of central thought processes traditionally

referred to as cognitive. These processes are under conscious control and they are not specific for some content domains. The contrast is thus no longer one between sensory-driven and concept-driven processes. Instead, one must now face up to a three-layered picture with sensory states, modular states, and belief states. The specific claim that is new with this notion of modularity is that there is a level of processing that is intermediate between sensations and full blown concepts. Such a level might be called a level of shallow object-recognition, shallow because not yet integrated in the network of real world knowledge. The example from language is helpful. Sensorially there are sounds, centrally there are meanings or concepts, while at the intermediate level there are linguistic objects or linguistic representations resulting from processing by linguistic structures, but not yet meaningful messages.

Modularity is thus not a claim to be taken lightly or to be given a relative interpretation, as if there were such a thing as being 'relatively (but not absolutely) domain-specific'. Of course, Fodor has somewhat approximately listed traits, as if presenting a full-blown checklist of identifying characteristics. But it is clear that characteristics like speed, impenetrability or pre-wiredness are found across the board off all information processes and ultimately they may therefore not be helpful to characterise a module. As we shall see below, the hypothetical mechanism of audio-visual spatial integration appears to score well on most of these traits, but we would argue that such compliance with superficial properties of modular processing does not make that mechanism a module in any interesting sense of the term. For example, it is a trivial fact that sensory processing is not under conceptual or doxastic influence of the kind usually assimilated with a subjects' cogitations. But that does not make it a module in the strong sense we want to defend here. A module then in the sense in which this notion may be needed and be useful for understanding cross-modal events, is a set or processing devices, a mechanisms for short, that is operational in a specific semantic environment. The study of modular processing is that of the processing abilities of a system qua functional architecture, in its biological sense. The latter is likely to exhibit some degree of species specificity as the example from language does bring out. In contrast, the study of modality of input concerns physical properties of the stimulus input. Principles, Gestalt ones like grouping, common fate, and others or stimulus properties like

signal intensity are general and found in vision just like in audition.

3. Audio-visual integration in the McGurk effect

The term 'McGurk effect', (MacDonald & McGurk, 1976) refers to the illusory percept in the presence of incongruent visual and auditory speech information. For example, when an auditory 'ba', is dubbed on the video of a speaker who says 'ga', listeners often report hearing the fused response 'da'. Most of the time, listeners do not report any conflict between the contradicting information and they report a syllable that is a compromise in terms of place of articulation features. The phenomenon has been replicated many times with different consonants and vowels as well. The subjective experience is always compelling: the subjects actually 'hears' the compromise syllable. Even when the listeners are told about the dubbing operations and instructed just to report the audio part of the stimulus, they cannot not avoid integrating the auditory and visual components of the speech stimulus. It implies that, even if the auditory speech is perfectly intelligible, listeners take into account the visual speech as well.

4. Audiovisual pairings in ventriloquism

The ventriloquist produces speech without visible facial movements and moves a puppet in synchrony. He thus creates a compelling impression that the speech comes from the mouth of the puppet. The illusion, which has come to be called 'the ventriloquist effect', occurs in other situations like amateur cinema set-ups, when the sound comes from a loudspeaker to one side of the screen and is nevertheless experienced as coming from the mouth of the actor or from other sound-producing objects. The effect was observed also by Stratton (1897) in the well-known experiment in which he wore for several days an optical device that reversed the visual field: he reports that when the sources of sounds were in sight, he often experienced them as coming from the displaced visual location of that source.

Jack & Thurlow (1973) have studied the conditions for ventriloquism in situations in which subjects were presented with speech through hidden loudspeakers and various candidate visible sources, and indicated through key-

pressing when they experienced perceptual fusion of the auditory and visual data. Synchronization of visible movements with peaks of speech intensity was found to be the main condition for the formation of the illusion.

An alternative way of studying ventriloquism consists of having the subject indicate, by pointing or by verbal report, the location of stimuli in one modality accompanied or not by conflicting stimuli in the other modality. In such selective attention situations, the usual finding has been that the reported location of the target stimuli is, on bimodal trials, shifted in the direction of the conflicting data, in comparison with the unimodal trials (Bertelson & Radeau, 1981, 1986, Pick, Warren & Hay, 1969; Radeau, 1992; Radeau & Bertelson, 1987; Thomas, 1941). This particular manifestation of ventriloquism has been called 'immediate cross-modal bias'. Biases of auditory location by visual data, and of visual location by auditory data have both been reported (Radeau & Bertelson, 1987), but the latter effect was generally of smaller amplitude, and has often been considered as non-existent.

The interaction between visual and auditory data in spatial perception are not limited to the situation involving speech and potential visible correlates. Strong impressions of common origin can also be created with, for instance, sound bursts coming from a hidden loudspeaker in one place and light pulses produced in synchrony in a different place, say 10 to 20 deg to one side (Choe, Welch, Gilford & Juola, 1975; Radeau & Bertelson, 1974).

5. Is the McGurk effect an instance of audio-visual grouping?

From the brief descriptions we gave of McGurk effects and of audiovisual pairings it might look as if both are closely related, either because the latter is a specific case of the former or vice versa. Let us first consider the possibility that McGurk effects would be a special case of audio-visual pairings. This would mean that the mechanism underlying for example a fusion response is not specific for speech processing and operates irrespective of whether the information is speech or not, obeying only certain constraints and requirements of contiguity in space/time such that parallel or concurrent extraction of information is possible. Such an approach seems to motivate the view long defended by Massaro and collaborators (see Massaro, 1987 for a full discussion). On this view there is continuous extraction

of featural information in the two modalities and continuous integration of the obtained featural information. At first sight this can account for the range of visual responses observed in the McGurk effect. The presence of the actual phonetic features determines the outcome of audiovisual percept. There is no need for the speech perception mechanism invoked to be modality specific. What is extracted from the two modalities is abstractly characterised featural information. In this sense, then, a model like Massaro's is not a model of bi-modal or modality specific perception at all. Neither the representations it postulates nor the processes invoked are specific for the modalities considered.

A critical question raised by the McGurk effect and noted early on by researchers concerns the requirement of a common metric between two very different input systems. If the speech perception system puts together, as it seems to do, speech information coming from audition and speech information coming from vision, this would require a common metric (whether this turns out to be a translation of one type of representations into the other or, alternatively, a third representation). For a model postulating modality-neutral feature extraction, the issue of a common metric hardly exists since presumably the abstract speech perceiver picks up the information directly in its own format. Before returning to this discussion, consider the alternative.

II. Content driven versus contiguity driven dynamics.

In the remaining of this paper we want to argue that recent evidence from various new sources now available may not be compatible with the notion of a generic processing system that operates over modality neutral representations putting together whatever information available in separate modalities. Instead, a model that combines content driven and contiguity driven processes seems to be required. An assumption of such a model would seem to be that representations are identified by a double signature, one derived from their content domain and one fixing their spatio-temporal identity.

1. The role of content as a motor of intersensorial integration.

There now seem to be clear examples of integration processes under the

control of content that are a puzzle for contiguity-based views. As noted when we discussed modularity, by content we here mean roughly the same as semantic domain or category, in the sense in which this term is used for example in cognitive neuropsychology to refer to category-specific disorders.

Cross-modal interactions will not occur whenever one of the inputs does not match the semantic domain of the module. For instance, Summerfield (1979) observed that auditory speech is not integrated with a Lissajou ring which corresponds to the centers and corners of the lips. Presumably, integration of speech only occurs if the input of both modalities is processed as speech, and a Lissajou figure does not fulfil this criterion. It thus seems that Gestalt-like and module-like criteria for integration are different: synchronicity and spatial adjacency are triggers for contiguity based processes, but they are not sufficient to guarantee module-like audio-visual speech integration.

A related example is the fact that a McGurk-like illusion does not occur when speech is paired with a written word instead of a speaking face. For example, Fowler and Dekle (1991) presented listeners with an auditory /ba/-/ga/ continuum and they simultaneously presented an orthographic 'BA' or 'GA'. Unlike in the McGurk effect, subjects never reported hearing the fused 'da', which is the common answer in the McGurk illusion when auditory /ba/ is combined with visual /ga/. Auditory speech and visual orthography are thus not integrated in the way auditory speech and visual speech are: orthography does not interact with speech processing. The subjective experience is also very different in these two cases, because when subjects sees letters instead of a face, it seems obvious that they are aware of the discrepancies between what they hear and read.

2. Content-based processes overrule contiguity related processes.

A further distinction between contiguity versus content-based interactions is that the former have their own signature, namely adaptation and recalibration. When two cues of the same perceptual parameter (e.g., depth, location etc) arrive at different values, adaptation, and later recalibration takes place. This situation is of course well demonstrated by the work of Bertelson on ventriloquism. However, adaptation and recalibration phenomena do not take place in the phonetic module.

Thus, in the McGurk-situation, conflicting phonetic cues from audition and vision are presented, and the outcome is a more or less optimal solution which fits both information sources (Massaro, 1987; Vroomen, 1992). There is, however, in the McGurk-situation no adaptation to the strange combination of auditory and visual cues, and even more important, so far nobody has reported any evidence for recalibration processes. It thus looks that conflicting cues in the phonetic module are solved differently if compared with conflicting cues in the spatial or temporal domain: the former do not lead to recalibration, the latter do. This strongly suggests that audio-visual speech and audio-visual pairing are two sides of two different coins: audio-visual speech is specific for the speech module, audio-visual pairing applies to audio-visual speech and to non-speech in a non-domain-specific way.

3. Dissociations between sensorial and modular impairments

Independent evidence for a focus on content and the need to consider commonality of content across differences in input modality is provided by neuropsychological impairments. Content based or modular impairments in the absence of sensorial disorders have been observed for example by Campbell (1993) and ourselves. We observed impairments in speech processing and in memory for speech input in developmental phonological dyslexics (de Gelder & Vroomen, 1995). This fact is intriguing because it concerns an impairment in the speech-processing domain across two very different input modalities. On the other hand, once the notion of modality-independent content is admitted, one is lead to expect an impairment across input channels. Thus, if poor readers suffer from phonological processing impairments, there is no reason to limit these to problems in the auditory modality. As a matter of fact, we have consistently found that in these subjects phonological impairments are also found when speech is lipread (de Gelder & Vroomen, 1988; de Gelder & Vroomen, 1995; Vroomen, 1992). Modular impairments thus ignore sensory modalities. The converse also seems to hold. An impairment may concern a domain of processing for which the carrier is a sensory modality that is also the carrier for other contents than the one impaired. For example, the face is a carrier of information for personal identity, for emotion expression, and for speech among others. But a face processing deficit does not

seem to lead to a visual speech perception deficit (Campbell, this volume; de Gelder, Vroomen & van der Heide, 1991; Gepner, de Gelder & de Schonen, submitted).

4. Within-domain modality effects

The examples given so far all tend to contrast modality-based processes with content or modularity-based ones. There is however also evidence suggesting that after or in the course of content-based processing, modality-specific markers remain present. A particularly clear area of evidence for modality-specificity within the speech module can be found in studies on short-term memory of heard and lipread speech (de Gelder & Vroomen, 1992; de Gelder & Vroomen, 1995). In immediate serial recall experiments, the subject is presented with a list of stimuli and asked to repeat that list. Recall is typically better for the last and penultimate items of the list, a phenomenon called recency. If at the end of such memory lists an item is added, the recency advantage of the last item is abolished, at least for the kinds of suffixes that share certain properties with the list items. For example, recency is left intact when a written word or a nonspeech sound is presented directly after the last item. In contrast, a spoken suffix strongly reduces the last item advantage for a spoken list, but it does so much less for a lipread list and vice versa. Based on such findings we have argued (de Gelder & Vroomen, 1994) that speech representations in immediate memory are still modality specific while being processed by an abstract mechanisms.

5. Experimental dissociations between content-based and contiguity-based processes.

The question of the relation between audiovisual pairings in ventriloquism versus the McGurk effect has rarely been considered in the literature, except for two unpublished dissertations (Fisher, 199*; Sharma, 199*). Massaro (1987), however, has argued that some of the data from ventriloquism studies brings support to his application of fuzzy logic theory to the interpretation of McGurk type effects. He was thus assuming that the two phenomena are based on at least partially common processes.

Experimental dissociations would consist of independent variables affecting one phenomenon and not the other one, or one more than the other. The candidate experimental variables were orientation of the speaker's face (upright vs. upside-down) and desynchronization of auditory from visual utterances. The results of Radeau and Bertelson (1977) led us to predict no effect of face orientation on ventriloquism. At the time the experiment was started, there were no data concerning effect of that variable on McGurk interference. Results showing such effects have become available in the meantime (Green, 1994; Jones & Mulhall, 1995). Regarding synchronization, there is abundant demonstration that is one of the most important conditions of ventriloquism.

In a recent series of experiments we investigated spatial (ventriloquism) and phonetic (McGurk) audio-visual illusions within the same experimental set-up (Bertelson, Vroomen & de Gelder, submitted). Speech fragments such as auditory /ana/ or /ama/ were delivered in various azimuthal locations while the face of a speaker was presented on a centrally located screen. The face was either articulating an utterance which could be congruent with the acoustic input or not, or staying still. After each presentation, the subject both pointed to the apparent location of the auditory source and identified the presented utterance. The situation thus made it possible to observe both ventriloquism and McGurk interference within the same trial. Ventriloquism was supposed to manifest itself through a displacement of the pointing responses to inputs from the lateral loudspeakers in the direction of the centrally situated video screen. This shift should occur when the face was seen speaking, not when it remained still. McGurk-interference could occur on trials with discordant auditory and visual speech, as the occurrence of a fusion response. The results showed that inverting the face (up-side down) had no effect on ventriloquism, while at the same time, the McGurk effect was attenuated. Thus, the attraction of the location of a sound towards a moving face was independent of the orientation of that face. In contrast, the McGurk illusion is depending on the orientation of the face: there were less fusions with an up-side down face if compared with upright orientation, suggesting that the two phenomena are subserved by different processing mechanisms.

6. Neuropsychological observations.

We recently observed that autistic subjects may be capable of normal auditory and visual speech processing, but they may not integrate the two input modalities. Instead, in the bimodal situation the information from the visual modality seems to be ignored and does not affect the percept (de Gelder, Vroomen & van der Heide, 1992; Gepner, de Gelder & de Schonen, in press). Such an absence of cross-modal integration does not seem to be due to an impairment in contiguity-based processes. Campbell, de Gelder and de Haan (submitted) found a LH advantage for recognition of mouth shapes corresponding to speech sounds in still photographs. This underscores the content or domain-based similarity of the two speech conveying modalities.

7. The relation between content-driven and contiguity-driven processes: hierarchy?

A hierarchic perceptual organization has several consequences worth considering. The first is that in the absence of perceptual grouping of the multi-modal input by Gestalt-like principles, no information will be passed on to the domain-specific modular processes. For instance, it seems clear that auditory and visual speech will not be integrated if they are to a-synchronous, and, although there are to our knowledge no studies which have addressed this issue, it might also be the case that McGurk-like fusions will be attenuated if the locations of the auditory and visual information are to disparate. The refusal of the perceptual system to integrate these cross-modal inputs is caused at a non-modular stage at which Gestalt principles operate. They should thus also be found with other stimuli and modalities. One would thus expect the Gestalt principles to operate across modalities and across specific domains.

III. A double signature of representations?

We now want to draw some suggestions from what has been reported so far. One theoretical suggestion concerns audio-visual conflict. It appears that in the McGurk illusion there is a resolution of conflicting information. There is only conflict once the module has operated successfully, that is, once linguistic information has been extracted from the two sensory modalities. The conflict between the input from

the two modalities would not occur if in each of the modalities the speech module had not detected linguistic information. Likewise it should be very clear that conflict and conflict resolution of the kind found in fusions is a process under modular control and not a sensorial interaction of the kind so well illustrated in the research by the Bertelson and collaborators. Nor is it conceptual integration that is the issue, in the sense of a resolution is under conscious control.

A second major consideration relevant to what we just said is that the ranges of the two phenomena are not co-existent. As mentioned above, ventriloquism, which concerns the perception of source location, occurs both with speech and with non-speech inputs. On the other hand, McGurk-type phenomena concern speech identity and appear to involve data with closure relation to respectively speech sounds and facial movements. Evidence in favor of the notion of separate processes comes from a study (Radeau & Bertelson, 1977) in which the subjects monitored speech coming from a centrally located loudspeaker and visual input presented on a laterally displaced video screen. In one condition, the visual data were the face of the speaker. In the other condition, they were diffuse light flashes synchronized with the intensity peaks of the speech. A same period of monitoring under the two conditions produced comparable recalibration effects on speech localization, measured through a pointing task. It seems unlikely that the diffuse light flashes which produced a full ventriloquism effect would affect identification of the auditory speech input.

A third remark is prompted by findings about modality-specificity of representations that are processed in content-based mechanisms. In other words, there seem to be modalities within modular mechanisms. Modularity has challenged the traditional contrast between sensory modalities and abstract concepts. There is another dimension to this challenge: the existence of modality-specificity within the module. The critical distinction needed to make room for modality in the modular mind is the one between pre-modular modality aspects versus post-modular modality aspects. This distinction may be critical for understanding the interplay between content-based and contiguity-based processes.

It follows from the previous remarks that it is potentially misleading to compare the McGurk-illusion with the case of audio-visual pairing, since the former

is most clearly a case of a domain-specific conflict. Evidently, there is only integration because there is successful modular processing, and this can only happen once the module has operated successfully, that is, once linguistic information has been extracted from the two sensory modalities. The integration between the input from the two modalities would not occur if in each of the modalities the speech module had not detected linguistic information. Likewise it should be very clear that the integration of the kind found in the McGurk situation is a process under modular control and not a sensorial interaction of the kind so well illustrated by the Bertelson/Radeau research. Nor is it a conceptual integration issue, or a conflict whose origin or resolution is under conceptual control.

Acknowledgements

Research reported in the present paper was partly supported by a grant from the Belgian Ministere de l'Education de la Communaute Francaise ('Action de recherche concertee' (Language processing in different modalities: comparative approaches). The research of Jean Vroomen has been made possible by a fellowship of the Royal Netherlands Academy of Arts and Sciences.

References

Bertelson, P., & Radeau, M (1976). Ventriloquism, sensory interaction and response bias: Remarks on the paper by Choe, Welch, Gilford and Juola. Perception and Psychophysics, 19, 531-535.

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. Perception and Psychophysics, 29, 578-584.

Choe, C. S., Welch, R. B., Gilford, R. M., & Juola, J. F. (1975). The "ventriloquist effect": Visual dominance or response bias? Perception and Psychophysics, 18, 55-60.

de Gelder, B. (in press). Modularity and logical cognitivism. In A. Clark (Ed.), Folk psychology and common sense. Oxford: Oxford University Press.

de Gelder, B., Bertelson, P., Vroomen, J., & Chen, H. C. (1995). Inter-language differences in the McGurk effects for Dutch and Cantonese listeners. Proceedings of the Fourth European Conference on Speech Communication and Technology, Madrid, pp. 1699-1702.

de Gelder, B., & Vroomen, J. (1988). Bimodal speech perception in young dyslexics. Paper presented at the 6th Australian Language and Speech Conference, Sydney.

de Gelder, B., & Vroomen, J. (1989). Models in the mind, modules on the lips. Behavioral and Brain Sciences, 124, 762-763.

de Gelder, B., & Vroomen, J. (1992). Abstract versus modality-specific memory representations. Memory and Cognition, 20, 533-538.

de Gelder, B. & Vroomen, J. (1994). Memory for consonants versus vowels in heard and lipread speech. Journal of Memory and Language, 33, 737-756.

de Gelder, B., & Vroomen, J. (1995). Phonological memory deficits in young and adult dyslexics. In B. de Gelder and J. Morais (Eds.), Language and literacy. Hillsdale, NJ: Lawrence Erlbaum Associates.

de Gelder, B., Vroomen, J., & van der Heide, L. (1991). Face recognition and lip-reading in autism. European Journal of Cognitive Psychology, 3, 69-86.
23-43.

Fodor, J. (1983). Modularity of mind. Cambridge, MA: MIT Press.

Fodor, J. (1984). Observation reconsidered. Philosophy of Science, 51.

Jack, C. E., & Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the "ventriloquism" effect. Perceptual and Motor Skills, 38, 976-979.

Massaro, D. W. (1987). Speech perception by ear and eye: a paradigm for psychological inquiry. Hillsdale, NJ: Lawrence Erlbaum Associates.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. Nature, 264, 746-748.

Pick, H. L., Jr., Warren, D. H., & Hay, J. C. (1969). Sensory conflict in judgements of spatial direction. Perception and Psychophysics, 6, 203-205.

Radeau, M. (1992). Cognitive impenetrability in auditory-visual interaction. In J. Alegria, D. Holender, J. Morais, & M. Radeau (Eds.), Analytic approaches to human cognition (pp. 41-55). Amsterdam: Elsevier Science Publishers.

Radeau, M., & Bertelson, P. (1974). The aftereffects of ventriloquism. Quarterly Journal of Experimental Psychology, 26, 63-71.

Radeau, M., & Bertelson, P. (1977). Adaptation to auditory-visual discordance and ventriloquism in semi-realistic situations. Perception and Psychophysics, 22, 137-146.

Radeau, M., & Bertelson, P. (1987). Auditory-visual interaction and the timing of inputs. Thomas (1941) revisited. Psychological Research, 49, 17-22.

Spelke, S. (1987). The development of intermodal perception. In P. Salapatek and L. Cohen (Eds.), Handbook of infant perception: From perception to cognition (pp. 233-273). Orlando: Academic Press.

Summerfield, Q. (1979). Use of visual information for phonetic perception. Phonetica, 36, 314-331.

Thomas, G. J. (1941). Experimental study of the influence of vision on sound localization. Journal of Experimental Psychology, 28, 167-177.

Vroomen, J. (1992). Hearing voices and seeing lips: investigations in the psychology of lipreading. Doctoral dissertation, Tilburg University.